Real-Time Registration of Video with Ultrasound using Stereo Disparity

Jihang Wang [1], Samantha Horvath [2], George Stetten [1,2,3], Mel Siegel [2], John Galeotti [2]

[1] Department of Biomedical Engineering, Carnegie Mellon University;
[2] Robotics Institute, Carnegie Mellon University;
[3] Department of Bioengineering, University of Pittsburgh;
http://www.vialab.org

ABSTRACT

Medical ultrasound typically deals with the interior of the patient, with the exterior left to the original medical imaging modality, direct human vision. For the human operator scanning the patient, the view of the external anatomy is essential for correctly locating the ultrasound probe on the body and making sense of the resulting ultrasound images in their proper anatomical context. The operator, after all, is not expected to perform the scan with his eyes shut. Over the past decade, our laboratory has developed a method of fusing these two information streams in the mind of the operator, the Sonic Flashlight, which uses a half silvered mirror and miniature display mounted on an ultrasound probe to produce a virtual image within the patient at its proper location. We are now interested in developing a similar data fusion approach within the ultrasound machine itself, by, in effect, giving vision to the transducer. Our embodiment of this concept consists of an ultrasound probe with two small video cameras mounted on it, with software capable of locating the surface of an ultrasound phantom using stereo disparity between the two video images. We report its first successful operation, demonstrating a 3D rendering of the phantom's surface with the ultrasound data superimposed at its correct relative location. Eventually, automated analysis of these registered data sets may permit the scanner and its associated computational apparatus to interpret the ultrasound data within its anatomical context, much as the human operator does today.

Keywords: ultrasound, stereo matching, disparity, computer vision, Sonic Flashlight.

1. INTRODUCTION

The term "augmented reality" has generally meant augmenting the reality of the human operator by the introduction of information beyond the normal powers of human perception. The added information often comes from imaging modalities such as ultrasound, MR, or CT. These imaging data may simply be displayed, or they may be further analyzed, for example, to provide the operator with graphical overlays of segmented structures. If the augmented reality system includes a video camera, analysis of the video data can further help augment the experience of the human operator. Finally, the analysis of the video stream and the other imaging data can be combined for computer analysis, as it is in the mind of the operator. Thus the "reality" being augmented can include not only the human's, but also that of the imaging scanner and associated computer algorithms.

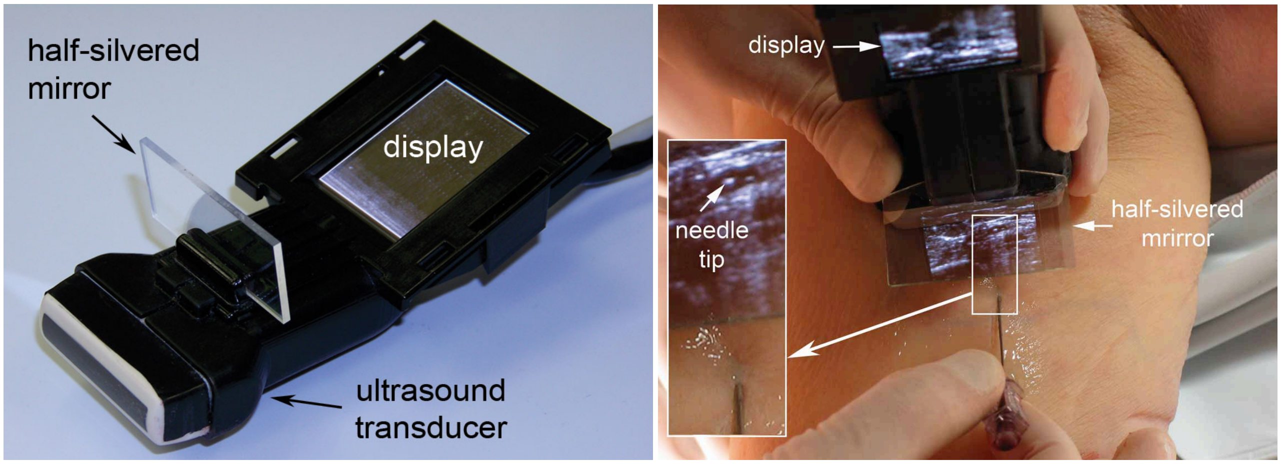

Figure 1. Sonic Flashlight device for viewing ultrasound in its actual location.

Our laboratory has developed a new augmented reality device, the Sonic Flashlight. The Sonic Flashlight merges an ultrasound image and the visual view of the operator using a half-silvered mirror and miniature display mounted directly on the ultrasound probe [1]. Figure 1 (left) shows the Sonic Flashlight and (right) the operator's point of view using it to guide insertion of a needle into a vein in the upper arm of a cadaver. The virtual image is shown magnified in the white box. The needle tip is visible as a bright spot within the dark cross section of the vein (adapted with permission from *Radiology* [2]). In the present research, we want to accomplish the same fusion of modalities within the scanner itself and its associated computational capabilities. This could eventually lead to automated analysis of the ultrasound data within its anatomical context, as derived from an ultrasound probe with its own visual input about the patient's exterior. We have previously accomplished a proof-of concept, using structured light from a pair of lasers and a single video camera, all mounted on the ultrasound probe, to determine the angle of the probe relative to the scanning surface [3]. We now do away with the lasers and determine the scanning surface using stereo matching between two video cameras mounted on the ultrasound probe.

2. USING STEREO CAMERAS TO DISPALY 3D SURFACE

2.1 Methods

The apparatus and gel phantom for the current experiment are shown in Figure 2. The photo is representative only, to show the constituent components. (The angle of the transducer would not actually produce an acceptable ultrasound image.) Twin video cameras (Minocam SD-008) mounted on the ultrasound probe (Diasus 15-22 MHz, from Dynamic Imaging) in a lightweight aluminum frame provide a view of the surface of the ultrasound phantom, upon which a sheet of tracing paper has been laid and saturated with gel. A checkerboard pattern has been printed on the tracing paper using an inkjet printer (a laserjet printer would melt the wax within the paper). The saturated tracing paper does not significantly interfere with the passage of ultrasound into the phantom.

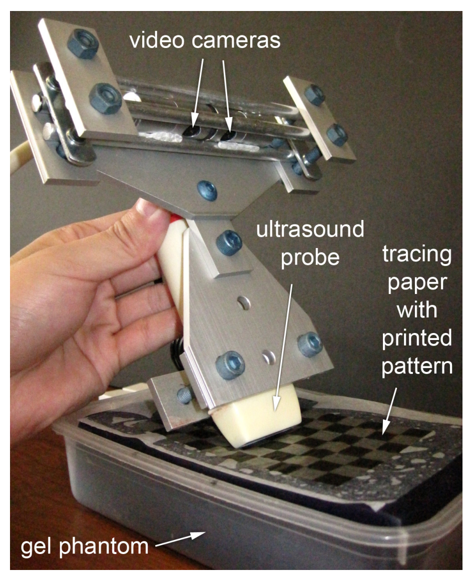

Figure 2. Apparatus and test phantom with printed surface.

The position of the phantom relative to the cameras is computed in real time by using *stereo vision* algorithms, wherein the separate viewpoints of cameras allow triangulation of the 3D coordinates of observed points. Each point thus identified on the phantom surface can be localized into the frame of reference of the cameras and, by extension, of the ultrasound probe itself. It is therefore possible to know both the shape and position of the viewed portion of the surface relative to the ultrasound probe in 3D in real time. Conversely, changes in position and orientation of the probe relative to the surface could also be computed. Although our specific application of stereo vision algorithms is novel, we are not developing new algorithms, but are instead relying on the algorithms freely available in OpenCV (Open Source Computer Vision Library), a widely used library of programming functions aimed at real-time computer vision, originally developed by Intel and now supported by Willow Garage.

2.2 Camera Calibration

When working with stereo vision, it is critical to know the spatial relationship between the cameras (the *extrinsic* parameters), and in many cases it is also important to get rid of the optical distortion due to imperfection of lens design and sensor placement (the *intrinsic* parameters). Since these parameters are fixed for our device, calibration need only be done once, prior to use. We employ Zhang's method [4] to automatically acquire the intrinsic parameters of each camera by showing it different orientations of a chessboard of known geometry [5], for which a simple contrast-based algorithm can locate the black-white intersections of the chessboard squares with sub-pixel resolution. The intersections appear in different locations from each camera's viewpoint, and these differences in pixel coordinates (the *disparity*) are used to compute the extrinsic parameters defining the spatial relationship between the cameras.

2.3 Stereo Rectification

Computation of stereo correspondence and disparity is more reliable and computationally tractable if the two image planes are exactly aligned. Unfortunately, it is impractical to build a perfectly aligned configuration, and so real stereo systems almost never have exactly coplanar, row-aligned imaging planes. *Rectification* compensates for this inevitable misalignment by using the extrinsic parameters to process the images in real time so that they are effectively coplanar and row-aligned. We use the Bouguet algorithm [6] for stereo rectification, which "minimizes the amount of change that projection produces for each of the two images while maximizing common viewing area." [5]

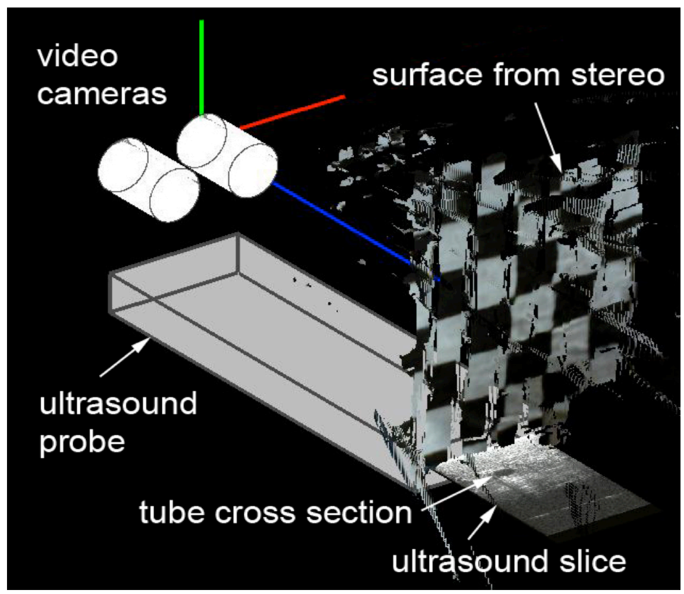

Figure 3. Real-time 3D simultaneous rendering of the gel phantom surface (from stereo), ultrasound, and probe/camera locations.

2.4 Stereo Matching and 3D Reprojection

Applying stereo vision to the phantom's surface requires some method of identifying and matching unique points that are visible to both cameras, analogous to the calibration section's identification of checkerboard intersections. Once such unique points, known as *correspondence points*, have been located in each of the images (i.e. put into correspondence), their disparity (horizontal pixel offset) may be readily computed within the two rectified images. It is then straightforward to calculate each correspondence point's 3D location based on its pixel coordinates, disparity, and the cameras' calibration data.

It is nontrivial to establish proper correspondence between points in an unconstrained scene. The enormous variety of surfaces in patients we might want to scan poses a critical problem. Luckily, there is also a wide range of stereo matching algorithms [7]. Since we require real-time 3D reconstruction, we use a fast and effective block-matching stereo algorithm that is similar to the one developed by Kurt Konolige [8]. It works by using small "sum of absolute difference" (SAD) windows to find matching points between the left and right stereo rectified images. Because our cameras are necessarily close to the skin's surface (about 120 mm), disparities of 110 pixels are not uncommon, despite our minimizing the baseline distance between the cameras. Robustly finding such large disparities is computationally difficult.

We achieve real-time performance only by using GPU acceleration, a new capability of OpenCV 2.2, which now offers NVIDIA CUDA implementations of select correspondence and 3D reprojection algorithms.

Reprojection of the surface is based on a similar triangles formula [5], which quickly calculates 3D coordinates for each correspondence point. These points are then textured with the video information and visualized with OpenGL in a 3D graphical environment. Since the location of the ultrasound data is known relative to the cameras, the ultrasound image can also be rendered (using 2D texture mapping) in the same 3D graphical environment in real time.

3. RESULTS

The results for the apparatus and phantom from Figure 2 are shown in Figure 3. The reprojection of the checkerboard pattern on the surface of the phantom has missing patches, since corresponding points in the stereo images were not found everywhere. However, a fairly consistent surface is evident. The ultrasound data is also projected in its correct location within the 3D coordinate space, and includes the cross section of a tube within the phantom. We use OpenGL to create a simulation ultrasound probe, which has the same size (height, width and length) as the experiment probe in order to indicate the moment-to-moment 3D location of the probe relative to the phantom surface. The OpenGL window can be rotated to see the surface and ultrasound image from any point of view. The particular point of view of the rendering as shown in Figure 3 is actually from *within* the gel phantom.

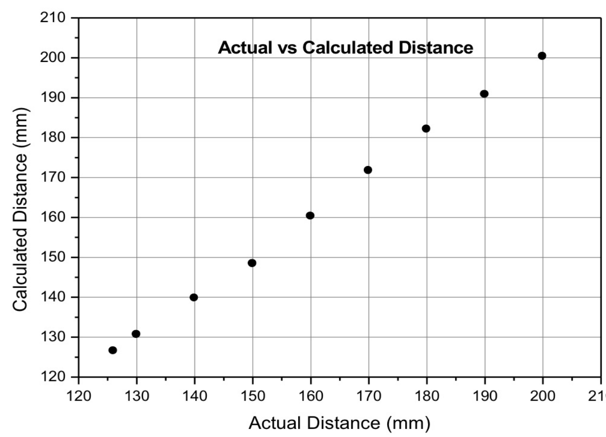

Figure 4. Validation of distance using stereo disparity.

Our apparatus can also report the computed distance from the cameras to the surface anywhere on the reprojected image. We tested this by putting the phantom at known distances relative to cameras pairs over an appropriate range. The results are shown in Figure 4 for nine separate trials. The RMS error is ±1.12 mm.

4. DISCUSSION

Clearly, a flat phantom is not a legitimate test for a clinically useful device. Human external anatomy consists of elastic tissue whose surfaces are not only curved but also deform and move relative to each other over an articulated skeleton. The simple checkerboard pattern is furthermore not a realistic representation of the natural markings on the surface of the skin. The matching algorithm we used finds only strongly matching (high-texture) points between the two images. Thus, in a richly populated scene such as a freckled and hairy arm, the majority of pixels might be attributed depth measurements, while in a very low-textured scene, such as hairless skin with uniform complexion, very few such points might be found. In these cases, our system could be aided by drawing dots or projecting structured light on the patient's skin. Lastly, our current reprojected surface is neither continuous nor as smooth as the actual surface of the phantom. Incorporation of a shape prior could improve this, as could data from previous video frames.

A number of other researchers have developed systems to merge ultrasound and visual data. The video information these systems capture can also be automatically analyzed using computer vision methods. Several have used head-mounted displays to overlay the ultrasound image on what the operator would see, as captured by video cameras [9][10]. Flaccavento, et al. described a system to track the location of an ultrasound probe in 3D space, using three stationary cameras that monitor patches adhered to the ultrasound probe. Since the cameras are stationary, this system restricts motion of the probe during the procedure and requires separate tracking of patient location [11]. Attaching the video camera directly to the ultrasound probe provides a simpler solution, as was proposed by Sauer and Khamene, who used it to permit graphical overlays on the video image showing a line of possible entry points for needle biopsy in the plane of the ultrasound scan [12]. Chan, et al., used stereo cameras mounted on the US probe with computer vision methods to determine needle location relative to probe [13].

More recently, Rafii-Tari, et al., attached passive optical markers to the skin along the patient's spine and used a camera mounted on an ultrasound probe to register the probe to the patient's anatomy, permitting accumulation of 3D ultrasound data and navigation within that data [14].

We aim to accomplish something along the lines of Rafii-Tari, but without requiring individually identifiable optical markers. Our goal is to use the surface of the patient directly as the visual coordinate frame for the computer, just as the human operator already does when scanning the patient with an ultrasound probe.

5. CONCULSION

Registering the surface of a target as computed from stereo video disparity with ultrasound data from within the target in real time is, to our knowledge, a new technique. We believe it has potentially wide-ranging utility in automated analysis of ultrasound by providing an anatomical context to the ultrasound data within the computational environment. We believe the research presented here represents important preliminary steps towards a clinically useful system, capable of operating in real time as the ultrasound probe is moved over the surface of the patient. In the mind of the human operator, this fusion is already an integral part of clinical practice, and thus it is an appropriate goal to provide the same fusion for analysis by automated systems.

6. FUTURE WORK

Recently, we have begun to track the visual texture on the surface of the patient/phantom in order to provide a solution when the probe position cannot be disambiguated based only on the external shape of the local anatomy and the ultrasound data. We do this by using 2D optical flow to establish temporal correspondences for the tracked 3D points across successive video frames. Since we also know the 3D location of each tracked point by stereo matching, we can then determine the correspondences between 2D image features and points on the 3D model based on the Iterative Closest Point algorithm [15]. In this way, we can track the full 3D movement of the probe relative to the patient's exterior.

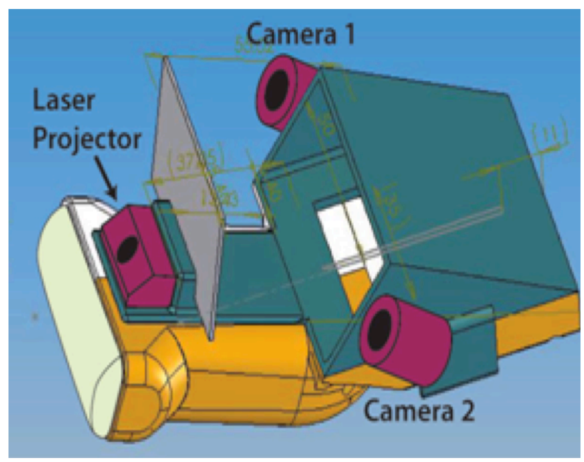

Figure 5. Plan for stereo cameras and a laser projector on the Sonic Flashlight.

We are also exploring the use of structured light to create rich features on the patient's skin to improve the quality of stereo matching. In addition to analyzing these images, we plan to transmit and record stereo video from the cameras, to be displayed on stereo-capable screens with LCD shutter glasses, overlaying the ultrasound data at its known location. This will allow direct fusion of the video and ultrasound data in the mind of the operator. We anticipate having to address the difference between the intra-camera distance and the intra-pupillary distance of the viewer in the graphical placement of the ultrasound image.

Ultimately, we want to merge these systems into a special model of the Sonic Flashlight (Figure 5) to provide a system that can combine the ultrasound and visual data in both the mind of the operator and in the software of the computer. We then intend to explore how the human and the machine can help each other to understand what each is seeing, for the benefit of the patient [14].

REFERENCES

1. Stetten, G., "System and Method for Location-Merging of Real-Time Tomographic Slice Images with Human Vision," U.S. Patent no. 6,599,247 (2003).
2. Chang, W., Amesur, N., Klatzky, R. Stetten, G., "The Sonic Flashlight Is Faster than Conventional Ultrasound Guidance to Learn and Use For Vascular Access on Phantoms," Radiology. Papers 241 (3), 771-779 (2006).
3. Horvath, S., Galeotti, J., Wang, B., Perich, M., Wang, J., Siegel, M., Vescovi, P., Stetten, G., "Towards an Ultrasound Probe with Vision: Structured Light to Determine Surface Orientation," Augmented Environments for Computer Assisted Interventions, Toronto, Canada (2011).
4. Zhang, Z., "Flexible Camera Calibration by Viewing a Plane from Unknown Orientations," Proc. ICCV, 666-673 (1999).
5. Bradski, G., Kaehler A., Learning OpenCV computer vision with the OpenCV library, O'REILLY, 1st Edition (2008).
6. Jean-Yves, "Camera Calibration Toolbox for Matlab," Rectification algorithm based on: Zhang, Z., "A flexible new technique for camera calibration," Proc. IEEE PAMI 22, 1330-1334 (2000).
7. Scharstein, D., Szeliski, R., "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," International Journal of Computer Vision (IJCV). Papers 47(1-3), 7-42 (2002).
8. Konolige, K., "Small vision system: Hardware and implementation," Proc. ISRR, 111-116 (1997).
9. State, A., Livingston, M., Garret, W., Hirota, G., Whitton, M., Pisano, E., Fuchs, H., "Technologies for Augmented Reality Systems: Realizing Ultrasound-Guided Needle Biopsies," Proc. SIGGRAPH 30, 439-446 (1996).
10. Sauer, F., Khamene, A., Bascle, B., Schimmang, L., Wenzel, F. & Vogt, S., "Augmented reality visualization of ultrasound images: System description, calibration, and features," Proc. ISAR, 30-39 (2001).
11. Flaccavento, G., Rohling, R., "Patient and Probe Tracking During Freehand Ultrasound," Medical Image Computing and Computer Assisted Intervention, Proc. LNCS 3217, 585-593 (2004).
12. Sauer, F., Khamene, A., "Video-assistance for ultrasound guided needle biopsy," U.S. Patent no. 6,612,991 (2003).
13. Chan, C., Lam, F., Rohling, R., "A needle tracking device for ultrasound guided percutaneous procedures," Ultrasound in Medicine & Biology. Papers 31(11), 1469-1483 (2005).
14. Rafii-Tari, H., Abolmaesumi, P., Rohling, R., "Panorama Ultrasound for Guiding Epidural Anesthesia: A Feasibility Study," Proc. IPCAI 6689, 179-189 (2011).
15. Zhang, Z., "Iterative Point Matching for Registration of Free-Form Curves and Surfaces," International Journal of Computer Vision (IJCV). Papers 13(2), 119-152 (1994).